

Dynamic Artificial Neural Networks for Centroid Prediction in Astronomy

S.J. Weddell & R.Y. Webb
University of Canterbury
Electrical & Computer Engineering
Christchurch, New Zealand
steve.weddell@canterbury.ac.nz

Abstract

Motivation for this research is the real-time restoration of faint astronomical images through turbulence over a large field-of-view. A simulation platform was developed to predict the centroid of a science object, convolved through multiple perturbation fields, and projected on to an image plane. Centroid data were selected from various source and target locations and used to train an artificial neural network to estimate centroids over a spatial grid, defined on the image plane. The capability of the network to learn to predict centroids over new target locations was assessed using a priori centroid data corresponding to selected grid locations. Various distortion fields were used in training and simulating the network including data collected from observation runs at a local observatory. Results from this work provide the basis for extensions and application to modal tomography.

1. Introduction

The real-time restoration of astronomical images, distorted by atmospheric turbulence, is the basis for the field of study known as adaptive optics (AO) [17]. Source objects, such as natural or artificial guide stars, are used in astronomy to provide a reference image for corrections to turbulence induced images of science objects, in real-time. To effectively compensate a science object by referencing a source object, the separation between both objects has until recently been spatially constrained. The largest angle of separation between such objects, where the turbulence induced by distortions can be considered identical, is referred to as the isoplanatic angle. The range of distortions that result from imaging the science object through a different portion of atmosphere is known as the anisoplanatic effect [14]. Substantial research has been directed towards minimising these effects [13].

More recently, tomography has been used for multi-

conjugate adaptive optics [4] (MCAO) to increase the anisoplanatic angle and hence reduce the dependency on nearby natural guide stars (NGSs), or carefully placed artificial sources known as laser guide stars (LGSs). For example, the suggestion of turbulence correction over the whole sky using multiple guide stars has been proposed [12]. Given such a proposal, limitations, in terms of both computational requirements and sensor technology, must first be addressed.

The objective of our research is to reduce computational overhead by extending the current spatio-temporal domain of turbulence prediction with the implementation of an artificial neural network (ANN). This paper will discuss the first stage of this work i.e., the establishment of a simulation platform and the development of an ANN that will provide short-term centroid prediction of a science object, given a series of perturbations that model atmospheric turbulence.

A brief background on modal tomography for adaptive optics and ANNs is given in Section 2. A discussion of our work follows in Sections 3 to 6. The results of our work are provided in Section 7 and this is followed by our conclusion and notes on future work in Section 8.

2 Background

A short discussion providing background information on two major disciplines, relevant to our research, is provided in this section. Section 2.1 discusses related work in adaptive optics and modal tomography. Section 2.2 provides an overview of ANNs and more specifically, dynamic artificial neural networks (DANNs).

2.1 Modal Tomography

The 3-D representation of an object using 2-D cross sectional *slices* is known as *tomography*. In astronomy, images from several guide stars are used to estimate 3-D perturbations by solving an inverse problem using methods similar to those employed in medical tomography [16]. A method

know as modal tomography has been developed by Ragazzoni [11] and applied over a large field [3] to correct for conical anisoplanatic distortion when laser guide stars (LGSs) are used as source objects. Three or more laser guide stars (LGSs), imaged through a telescope's *meta-pupil*, can be used to estimate aberrations that can be described by a series of orthonormal basis functions, known as Zernike polynomials [10]. The orthogonality of Zernike polynomials is expressed as,

$$\int_0^{\Upsilon} \int_0^{2\pi} F(\rho) Z_j(\rho, \theta) Z_k(\rho, \theta) \rho d\rho d\theta = \delta_{jk}, \quad (1)$$

where Υ is the radius of a circular pupil, $F(\rho)$ is the aperture weighting function, $Z_j(\rho, \theta)$ is the j th Zernike polynomial, and δ_{jk} is the Kronecker delta.

Ragazzoni [11] showed that given a wavefront, comprising an ensemble of Zernike polynomials of order Q and over a large pupil $P^Q(x, y)$, a smaller circular portion can be described by another Zernike ensemble $P'^Q(x, y)$, inside the larger pupil. Thus, a modal approach using Zernike polynomials can be employed as an alternative to the original method known as zonal tomography [15].

Modal tomography is not limited to LGSs. Unlike LGSs the extraction of low-order Zernike terms using wavefront sensing on natural guide stars (NGS) can include *Tip* and *Tilt* terms, Z_2 & Z_3 respectively. As a result, low order aberrations can be determined over the image of a science object by utilising the images of multiple NGS sources.

In the utilisation of modal tomography, the wavefront of a distorting medium can be determined using several natural or artificial source objects. For example, aberrations, comprising a set of Zernike terms, can be taken from several NGSs or artificial LGSs and used in conjunction with knowledge of the turbulence distribution, the height of the turbulent layers h , and the air refractive index fluctuations $C_n^2(h)$, to correct for an extended isoplanatic angle between multiple source objects and a science object. The isoplanatic angle, θ_0 , is given by,

$$\theta_0 = \left[2.91 k^2 \sec^8/3 \zeta \int_{Path} C_n^2(h) h^{5/3} dh \right]^{3/5}, \quad (2)$$

where $k = 2\pi/\lambda$, ζ is the zenith angle, C_n^2 is the structure constant of the index of refraction, and h is the altitude of the turbulence.

Modal tomography is used in adaptive optics to extend the isoplanatic *patch* surrounding a science object by imaging multiple source objects over a large field. The set of linear equations originally used to correct for conical anisoplanatism [11] is given as,

$$L_i = \sum_{j=1}^M L_{i,j} = \sum_{j=1}^M A_{ij} W_j, \quad (3)$$

$$W_T = \sum_{j=1}^M W_{Tj} = \sum_{j=1}^M T_j W_j, \quad (4)$$

where L_i is a set of natural or artificial sources, each element comprising a set of Zernike polynomials $Z_2 \dots Z_{Q+2}$, $A_{i,j}$ is the *interaction matrix* defining coefficients for each source over M turbulent layers, W_j is the meta-pupil of the field for M turbulent layers, W_T is the sum of all perturbations for M layers, adversely affecting the image of the science object.

Our aim is to predict W_T over Zernike terms Z_2 to Z_6 from an N -step time series, using data *a priori* from a set of source objects. In this paper, we discuss the short-term predictability of the second and third order coefficients, namely, the *tip* and *tilt* terms, Z_2 & Z_3 . Each term is predicted using centroid data over each source object through simulation.

The characteristics of turbulence can be learned using an ANN, given centroid data from multiple source objects. The *interaction matrix* defined in Equation 3 is unique for each observation site. In addition, the atmospheric seeing conditions, fluctuations in the refractive index of turbulence, and the altitude of the turbulence require frequent updates to minimise errors.

Current solutions to Equations 3 & 4 require complex matrix inversions and application of singular value decomposition (SVD). Such techniques are computationally intensive. Machine learning systems have been used for AO [9] and we propose an ANN as an alternative to matrix inversion [16]. Computationally efficient ANNs have been developed that employ the discrete cosine transform (DCT) for feature extraction resulting in a dimensional reduction by up to two orders of magnitude [2].

2.2 Artificial Neural Networks

Artificial neural networks (ANNs) are inspired by biological neurons and their interconnections that form the structure of the brain. ANNs can be defined as a network of nodal structures each comprising one or more input and output nodes, a weighted matrix of interconnections providing communication between each input and output, and an activation function that ensures each output is bounded.

For temporal analysis the implementation of a dynamic network is required. Unlike static networks, a dynamic network supports feedback and signal delays. A recurrent network supports a feedback structure. A dynamic feed-forward ANN does not support feedback but does support a tapped delay line for input data. The architecture of a dynamic feed-forward network used for this research is shown in Figure 1.

With reference to Figure 1, a dynamic feedforward ANN architecture is used for predicting the next time sampled in-

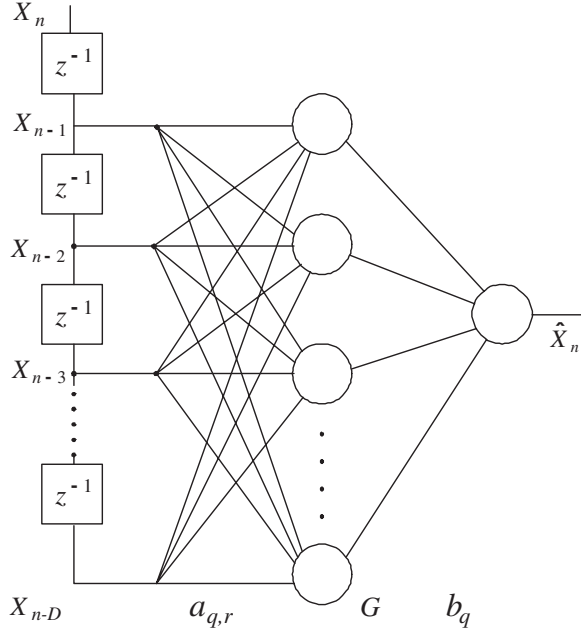


Figure 1. A Dynamic Feed-forward ANN

put, \hat{X}_n , given D time delays. The architecture is defined as,

$$\hat{X}_n = \sum_{q=1}^G b_q \tanh(a_{q,0} + \sum_{r=1}^D a_{q,r} X_{n-r}), \quad (5)$$

where, G is the number of artificial neurons and D is the dimension of the network comprising the number of tapped delays representing time-series data $X_{n-1} \dots X_{n-1+D}$, a & b are the input and output coefficients respectively, and q & r are indices used to represent each hidden node and time delay coefficient, respectively.

Back-propagation [6] is used to train the network. During training a series of targets is presented to the network and the coefficients adjusted to minimise the mean squared error, $E = (\hat{X}_n - X_n)^2$.

Given our current implementation, recurrent ANNs, such as Elman networks [6], generally perform better than feed-forward networks when the number of past data points is unknown [7]. Each hidden node within an Elman network supports a feedback connection with time delay. Several variations on feedback configurations have been used for time-series prediction [8]. To improve the accuracy of predicting more than one time-step into the future, a dynamic structure that supports global feedback is required.

A genetic algorithm has been developed to optimise performance in predicting centroids one sample period into the future [5]. Our work aims to extend the prediction of wave-front maps, initially with low-order aberrations such

as tip/tilt terms, over a large field-of-view.

3 Problem Statement

A data ensemble from the simulation of multiple source and target objects, projected on to a 2-D image space, Ω , and distorted by V perturbation sources, was obtained for analysis by an ANN. The ensemble comprised N time-series data, sampled at T_s second intervals. Centroid data from L sources $S_{x(0...L)}$ & $S_{y(0...L)}$ at spatial coordinates $h(x_0, y_0) \dots h(x_L, y_L)$, formed part the ensemble. These data sources were combined with centroid data from H target references $R_{x(0...H)}$ & $R_{y(0...H)}$, including a spatial coordinate for each reference, to train an ANN. We are interested in determining the minimum number of source objects, L , required to provide minimum variance, σ , over a line segment h_1, h_2 . In 2-D this is represented by the area, K .

A graphical simulation tool was used to represent the preceding problem statement and this was developed and implemented in C++. Simulation tools for the study of imaging through atmospheric turbulence have been used [1], however the development of a simplified model provided additional flexibility to allow us to explore the underlying behaviour of the system.

The simulation functions supported by our model, referred to as *Atmo*, currently include:

- Diffraction limited imaging: All source objects are imaged as point spread functions (PSFs) using Fraunhofer diffraction over a space invariant image plane. Changing telescope parameters such as aperture size, focal length, or the dimensions of the imaging plane, will affect the diffracted image.
- Source characteristics: Source objects can be created and parameters such as the irradiance and position can be defined over the image plane.
- Simulation parameters: The simulation period can be modified and the overall intensity of the field can be varied to model the frame rate of the imaging device and the CCD integration time respectively.
- Perturbation & noise parameters: The ability to select perturbation sources and emulate noise sources, such as low irradiance objects and CCD read-out noise, are supported.

A screen view of *Atmo* is shown in Figure 2. Further discussion is provided in Section 6.

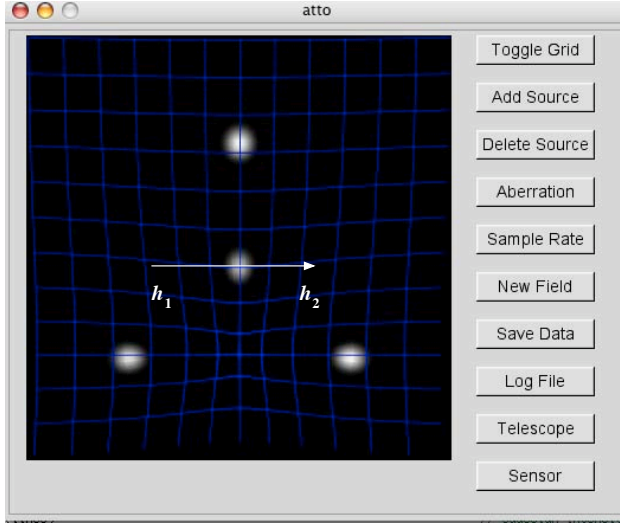


Figure 2. The Atmo Simulation Platform

4 Centroid Prediction

The collection of centroid data from multiple source objects was required to predict low-order aberrations, such as *tip* and *tilt* measurements, displayed in the image of a faint science object, distorted by turbulence. The Zernike terms Z_2 and Z_3 , discussed in Section 2.1, can be used to describe *tip* and *tilt* aberrations over a unit circle, represented by a mirror used in an astronomical telescope.

Given a region of interest (ROI), Γ , the centroid of Γ is a point within the region that is equally balanced. To obtain centroid estimates of the L initial positions of source objects, and H training objects over the image space, Ω , $L \times H$ target references were used to train an ANN implemented using Matlab's Neural Network Toolbox. A ROI was centred over the initial position of each source and target object prior to simulation. The centroids for each source and training object were then calculated for each simulation step. The resulting data were saved to a data file for further processing. The centroid equations are given as:

$$\hat{x} = \frac{\sum x_k(I_k - S)}{\sum I_k - S}, \quad (6)$$

$$\hat{y} = \frac{\sum y_k(I_k - S)}{\sum I_k - S}, \quad (7)$$

where x_k and y_k are the relative positions in the x or y direction of pixel locations with offset k from a fixed reference $\Gamma(x_0, y_0)$, I_k is the pixel intensity at (x_k, y_k) , and S is the mean pixel intensity representing the sky background radiation over the image plane, Ω .

The resulting data ensemble, comprising spatio-temporal data for L sources, were used to train and simulate the ANN. Gaussian and Poisson noise, following the discussion in Section 3, were added to centroid data. An example of three centroid pairs comprising 166 data samples, is shown in Figure 3.

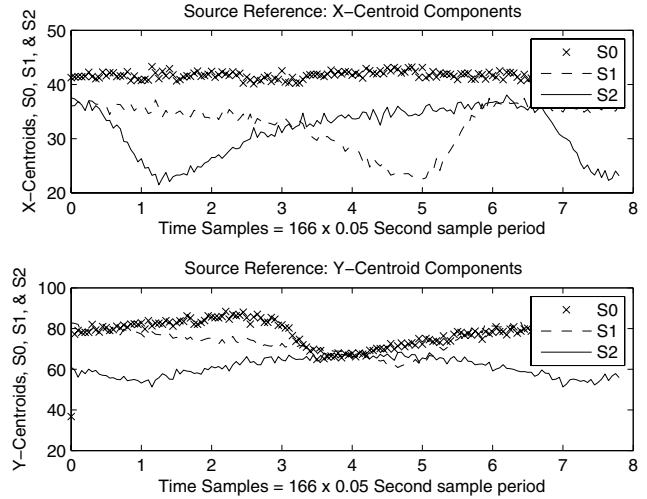


Figure 3. Source Centroid Data

5 Performance Considerations

An overview of ANN architectures and preliminary design considerations is discussed in Section 2.2. Based on these considerations a dynamic feed-forward network, similar to that shown in Figure 1, was used for this application. Network performance was based on the optimisation of two parameters, the network dimension, D , and the number of hidden nodes, G .

Several data vectors were presented to the network and the performance, in terms of the MSE on a trained data series, was monitored. If insufficient dimension D were applied to the network, the attractor of the system was projected into a space of insufficient dimension. If excessive dimensions were used then poor performance due to unnecessary complexity resulted.

In addition, the output mapping function lead to wide variations in output vectors and produced inconsistent results. An optimal embedding dimension can be determined using singular-value analysis and evaluating the eigenvalues of the covariance matrix. By varying the input dimension and examining each set of eigenvalues, the point where these values reach a minimum determines the optimum embedded size. Optimal performance was achieved for $D = 8$.

and $G = 100$.

6 Training Requirements

Back-propagation was used to train the ANN with known *a priori* centroid data. Each set of centroid data had an associated coordinate data. The resulting time series used for training comprised data from two training objects taken over 166 samples. Two training sequences were performed, the first used three source objects, the second used two.

The first training sequence was completed using sources $S_{x,y(0..2)}$, comprising all three source objects shown in Figure 2; the second training sequence was performed using two sources $S_{x,y(0..1)}$, where only the upper and lower left source shown in Figure 2 were used. Each source was equally spaced with an angular separation equivalent to 2-arcminutes.

Three data sets were collated from multiple simulations and these included source, training, and validation data. Data from spatial coordinates, $R_{x,y(0..1)}$, over a line segment h_1 and h_2 in Figure 2, were used to train the network.

The *generalisation ability* is a measure of how well the network can deal with inputs outside the training set. Over-fitted data can result in high errors on untrained input patterns but low errors on trained data. A network that has over-fitted the data suggests poor generalisation.

Randomising the input locations for training is expected to improve training times and the network's ability to predict new input sequences. However, this was not undertaken in the current investigation.

7 Results

The design of the ANN architecture requires optimisation in order to minimise the cost function over an extended spatial field.

The cost function comprised the mean squared error (MSE) and was measured as a function of the spatial distance between two source references, $R_{1,2}(x, y)$, over a line segment between normalised x and y grid locations, where $h_1 = (0.4, 0.5)$, and $h_2 = (0.6, 0.5)$. The length of 0.2 on this line segment is equal to an angular separation of 1-arcminute.

Forty, evenly spaced training points on the line segment between h_1 to h_2 shown in Figure 2 were considered potential locations of the science object and these locations were used to test the network's ability to predict centroids one time-step into the future, \hat{X}_n . Minimising the cost function over this segment, utilising the minimum number of source references, was our primary interest. Once the network was trained the simulation was repeated for each probe point as

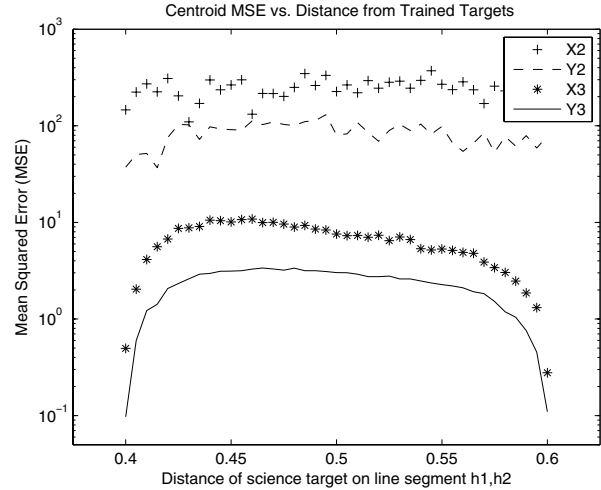


Figure 4. MSE between target references

the distance from the reference source in the x -direction was increased. The graph in Figure 4 shows the MSE for two data sets, both sets are a function of the Euclidian distance between each trained end-point on the line segment, h_1, h_2 .

The results in Figure 4 show the MSE pair, X3 and Y3 for the prediction of centroids of a science object using three source objects, $S_{x(0..2)}$ & $S_{y(0..2)}$. The second MSE pair X2, and Y2, show the results of centroid prediction using two source objects, $S_{x(0,1)}$ & $S_{y(0,1)}$. These results show a significant improvement in the network's ability to predict centroids over a 1-D line segment with the addition of a third source object.

The sharp increase in MSE over a simulated separation of 10-arcseconds between each target's trained end-point shows a low MSE on the trained end locations. However, the constant response over the majority of the line segments suggest the network's ability to deal with unforeseen inputs. As a result, the *generalisation ability* of the network was considered reasonable however further analysis will be required on randomly trained data to assess the network's ability to maintain a lower MSE over a longer path length and for extensions to regions over the image plane.

8 Conclusion & Future Work

In this paper we propose an alternative method for solving the *interaction matrix*, required in modal tomography for the prediction of turbulence over science objects by using a DANN without the need to apply matrix inversion. A simulation platform has been implemented and centroid data over a science object has been predicted by a DANN

using multiple sources over a large field-of-view.

Currently, we are optimising the DANN architecture to improve the MSE in 1-D and this will be extended to a 2-D region over 2 arcminutes. Computational capabilities of our network will be monitored and compared with the computational requirements to solve the inverse problem used in modal tomography.

We also plan to explore the application of real-time recurrent learning [6] and extend centroid predictions over a 5-arcminute spatial field. Optimisation of the network architecture to provide centroid predictions in real-time and include higher order aberrations over multiple time-steps into the future are considerations for future work.

References

- [1] M. Carillet, C. Verinaud, B. Femenia, A. Riccardi, and L. Fini. Modelling astronomical adaptive optics - the software package CAOS. *Royal Astronomical Society*, 365:1263–1275, 2005.
- [2] G. Chundi, M. Lloyd-Hart, and M. Sundareshan. Training multilayer perceptron and radial basis function neural networks for wavefront sensing and restoration of turbulence-degraded imagery. *Proceedings of the IEEE International Joint Conference on Neural Networks*, 3:2117–2122, 2004.
- [3] T. Fusco, J.-M. Conan, L. Michau, and G. Rousset. Efficient phase estimation for large-field-of-view adaptive optics. *Optics Letters*, 24(21):1472–1474, 1999.
- [4] T. Fusco, J.-M. Conan, L. Michau, G. Rousset, and F. Assemat. Multi-conjugate adaptive optics: Comparison of phase reconstructions for large field of view. *Proceedings of SPIE - The International Society for Optical Engineering*, 4167:168–179, 2001.
- [5] P. Gallant and J. Aitken. Genetic algorithm design of complexity-controlled time-series predictors. *IEEE XIII Workshop on Neural Networks for Signal Processing*, pages 769–778, Sept 2003.
- [6] J. Hertz, A. Krogh, and R. Palmer. *Introduction to the Theory of Neural Computation*. Westview Press., 1991.
- [7] T. M. Inc. Technical solutions. *The Mathworks Inc.*, (1-15KT9), July 2006.
- [8] D. Mandic and J. Chambers. *Recurrent Neural Networks for Prediction - Learning Algorithms, Architectures and Stability*. John Wiley & Sons Ltd., 2002.
- [9] P. McGuire, D. Sandler, M. Lloyd-Hart, and T. Rhoadarmer. Adaptive optics: Neural network wavefront sensing, reconstruction, and prediction. *Scientific Applications of Neural Nets, Proceedings of the 194th W. E. Heracus Seminar, 1998*. Edited by J.W. Clark, T Lindenau, and M.L. Ristig, Springer-Verlag Publishers, page 97, 1999.
- [10] R. Noll. Zernike polynomials and atmospheric turbulence. *Journal of the Optical Society of America*, 66(3), 1976.
- [11] R. Ragazzoni, E. Marchetti, and F. Rigaut. Modal tomography for adaptive optics. *Letter to the Editor, Astronomy and Astrophysics*, 342(L53-L56), 1999.
- [12] R. Ragazzoni, E. Marchetti, and G. Valente. Adaptive optics corrections available for the whole sky. *Nature Letters*, 403(6765):54–56, 1999.
- [13] F. Roddier. *Adaptive Optics in Astronomy*. Cambridge University Press, 1990.
- [14] M. Roggermann and R. Welsh. *Imaging Through Turbulence*. CRC Press Inc., 1996.
- [15] M. Tallon and R. Foy. Adaptive telescope with laser probe: isolanatism and the cone effect. *Astronomy and Astrophysics*, 235:549, 1990.
- [16] A. Tokovinin, M. L. Louarn, E. Viard, N. Hubin, and R. Conan. Optimised modal tomograph in adaptive optics. *Astronomy & Astrophysics*, 378:710–721, 2001.
- [17] R. Tyson. *Introduction to Adaptive Optics*. SPIE Press, 2000.